

# **다변량분산분석(MANOVA)**

# [Review] One-way ANOVA

- 분산분석

- 비계량(non-metric) 독립변수와 계량(metric) 종속변수 사이의 관계나 영향을 분석하는 방법
- 독립변수 : 범주형이며, 변수의 수로 분산분석 결정
  - 변수형태 : 요인(factor), 처리(treatment), 집단(group)
  - 범주(수준) : 요인을 구분하는 분류 기준을 의미
- 종속변수의 수에 따라 분류
  - 일변량 분산분석(ANOVA) : 종속변수가 하나인 경우
  - 다변량 분산분석(MANOVA) : 종속변수가 둘 이상인 경우
- 가정 :  $Y_{ij} \sim \text{iid } N(\mu_i, \sigma^2) \ (i = 1, 2, \dots, a)$ 
  - 각 범주(수준)마다 동일한 분산을 갖는 정규모집단에서 추출된 확률표본으로 가정
  - 각 범주(수준)의 모집단 분포는 서로 독립

- 자료구조 : 1원배치 분산분석

- a 개의 수준, 각 수준마다  $n_i$  개 표본(관측자료) 가정

| 수준         | Factor A                         |                                  |     |                                  |
|------------|----------------------------------|----------------------------------|-----|----------------------------------|
|            | 1                                | 2                                | ... | a                                |
| 자료<br>(표본) | $Y_{11}$<br>$Y_{12}$<br>$\vdots$ | $Y_{21}$<br>$Y_{22}$<br>$\vdots$ |     | $Y_{a1}$<br>$Y_{a2}$<br>$\vdots$ |
|            | $Y_{1, n_1}$                     | $Y_{2, n_2}$                     |     | $Y_{a, n_a}$                     |
| 표본평균       | $\bar{Y}_1$                      | $\bar{Y}_2$                      |     | $\bar{Y}_a$                      |
| 전체 평균      | $\bar{Y}$                        |                                  |     |                                  |

모집단 : iid  $N(\mu_i, \sigma^2)$

$$\bar{Y} = \frac{\sum_{i=1}^a n_i \bar{Y}_i}{N} = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^a n_i}$$

- 분산분석 모형

- i 수준의 모평균을  $\mu_i$  로 표현할 때,

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (i = 1, 2, \dots, a ; j = 1, 2, \dots, n_i) \text{ where } \varepsilon_{ij} \sim N(0, \sigma^2)$$

$$= \mu + (\mu_i - \mu) + \varepsilon_{ij} \quad \text{where} \quad \mu = \frac{\sum_{i=1}^a n_i \mu_i}{N} = \frac{\sum_{i=1}^a \sum_{j=1}^{N_i} Y_{ij}}{\sum_{i=1}^a N_i}$$

$$= \mu + \alpha_i + \varepsilon_{ij} \quad \text{where} \quad \alpha_i = \mu_i - \mu$$

- 요인 A의 주효과(main effect)의 추정

$$\alpha_i = \mu_i - \mu \Rightarrow \hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{Y}_i - \bar{Y}$$

- 변동분해

$$\sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2 = \sum_{i=1}^a n_i (\bar{Y}_i - \bar{Y})^2 + \sum_{i=1}^a \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

$$\Leftrightarrow \text{SST} = \text{SSA} + \text{SSE}$$

$$\Rightarrow \text{df}_T = \text{df}_A + \text{df}_E \Leftrightarrow N - 1 = (a - 1) + (N - a)$$

- 분산분석표

| Source | SS  | df    | MS  | F                 |
|--------|-----|-------|-----|-------------------|
| Model  | SSA | a - 1 | MSA | $\frac{MSA}{MSE}$ |
| Error  | SSE | N - a | MSE |                   |
| Total  | SST | N - 1 |     |                   |

where  $MSA = \frac{SSA}{a - 1}$  and  $MSE = \frac{SSE}{N - a}$

- 검정 과정/F-검정

① 가설설정 : 요인 수준간 평균 차이(=효과 차이) 비교

$$H_0 : \text{All } \mu_i \text{ are equal (or All } \alpha_i = 0) \text{ vs } H_1 : \text{Not } H_0$$

② 검정통계량과 분포  $F_o = \frac{\text{수준간 변동}}{\text{수준내 변동}} = \frac{MSA}{MSE} \sim F(a - 1, N - a) \text{ under } H_0$

③ p-값 계산과 판단

$$p - \text{value} = \Pr \{F(a - 1, N - a) > F_o \mid H_0\} \Rightarrow \text{if } p - \text{value} < \alpha, H_0 \text{ is rejected}$$

④ 사후분석(=다중비교)

- 귀무가설이 기각되면, 구체적으로 어느 수준에서 차이가 있는 지를 분석하는 과정으로 평균이 동일한 수준(그룹)을 판단
- 방법 : LSD, Tukey, Scheffe, SNK, Duncan/Bonffernoi etc

# MANOVA : 일원배치 다변량 분산분석

- 다변량분산분석의 목적
  - [Review] 일변량 분산분석 : 요인수준에 대한 종속변수의 모평균 차이 검정
  - 둘 이상의 종속변수에 대한 요인수준간 **모평균벡터의 차이를 검정**하는 것이 목적
- 자료 구조 :  $p$ -차원 벡터  $\mathbf{Y}'_{ij} = [Y_{ij1} \ Y_{ij2} \ \dots \ Y_{ijp}]$  로 가정

| 수준               | Factor A               |                        |     |                        |
|------------------|------------------------|------------------------|-----|------------------------|
|                  | 1                      | 2                      | ... | a                      |
| 자료<br>벡터<br>(표본) | $\mathbf{Y}'_{11}$     | $\mathbf{Y}'_{21}$     |     | $\mathbf{Y}'_{a1}$     |
|                  | $\mathbf{Y}'_{12}$     | $\mathbf{Y}'_{22}$     |     | $\mathbf{Y}'_{a2}$     |
|                  | $\vdots$               | $\vdots$               | ... | $\vdots$               |
|                  | $\mathbf{Y}'_{1, n_1}$ | $\mathbf{Y}'_{2, n_2}$ |     | $\mathbf{Y}'_{a, n_a}$ |

- 가정 : 일변량 분산분석과 거의 동일
  - 요인의 각 수준별 자료는 동일한 공분산행렬을 따르는 다변량 정규모집단에서 랜덤하게 추출된 자료로 가정
  - 인자의 각 수준에 대한 정규모집단은 서로 독립으로 가정

$$\mathbf{Y}_{ij} \sim \text{iid } N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \text{ where } i = 1, 2, \dots, a$$

- 모형 설정 :
  - 일변량분산분석의 형식과 동일하나 변수를 벡터로 표현하는 차이가 존재

$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{ij} \text{ where } \boldsymbol{\varepsilon}_{ij} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \boldsymbol{\alpha}_i = \boldsymbol{\mu}_i - \boldsymbol{\mu}$$

$$\mathbf{Y}_{ij} = \begin{bmatrix} \mathbf{Y}_{ij1} \\ \mathbf{Y}_{ij2} \\ \vdots \\ \mathbf{Y}_{ijp} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{ij1} \\ \boldsymbol{\varepsilon}_{ij2} \\ \vdots \\ \boldsymbol{\varepsilon}_{ijp} \end{bmatrix}$$

## • 변동분해

$$\sum_{i=1}^a \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}})' = \sum_{i=1}^a \mathbf{n}_i (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})' + \sum_{i=1}^a \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)'$$

$\Leftrightarrow \mathbf{T}_{p \times p} = \mathbf{B}_{p \times p} + \mathbf{W}_{p \times p} = \text{수준간 (Between) 변동} + \text{수준내 (Within) 변동}$

– 수준내 변동  $\mathbf{W}_{p \times p}$  의 의미

$$\mathbf{W}_{p \times p} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)' = \sum_{i=1}^a (\mathbf{n}_i - 1) \mathbf{S}_i$$

————— 각 수준의 공분산행렬이 동일하다는 가정에서의 추정된 합동 표본공분산행렬 —————  $\rightarrow \mathbf{S}_p = \frac{\mathbf{W}_{p \times p}}{N - a}$

## • 분산분석표

| Source  | 제곱합과 교차곱 행렬(SS)   | df      |
|---------|---|---------|
| Between | $\mathbf{B}_{p \times p} = \sum_{i=1}^a \mathbf{n}_i (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})(\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})' [= \mathbf{H} \text{ in SAS}]$   | $a - 1$ |
| Within  | $\mathbf{W}_{p \times p} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)' [= \mathbf{E} \text{ in SAS}]$ | $N - a$ |
| Total   | $\mathbf{T}_{p \times p} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}})'$                                   | $N - 1$ |



## • 다변량 분산분석의 검정 과정

### ① 가설검정 : 각 수준에 대한 모평균벡터가 모두 동일함을 검정

$H_0$  : All  $\mu_i$  are equal (or All  $\alpha_i = 0$ ) vs  $H_1$  : Not  $H_0$

$$\text{cf. } H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{a1} \\ \mu_{a2} \\ \vdots \\ \mu_{ap} \end{bmatrix}$$

### ② 검정통계량과 분포 $\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}$ = Wilk's $\lambda \sim F - dist$ . (작을수록 기각)

• Wilk's Lambda의 분포

| p<br>(차원) | a<br>(수준수) | 다변량 정규자료에 대한 근사 분포   |
|-----------|------------|--|
| 1         | $\geq 2$   | $\Lambda_o = \left( \frac{N-a}{a-1} \right) \left( \frac{1-\Lambda^*}{\Lambda^*} \right) \sim F(a-1, N-a)$                         |
| 2         | $\geq 2$   | $\Lambda_o = \left( \frac{N-a-1}{a-1} \right) \left( \frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F(2(a-1), 2(N-a-1))$ |
| $\geq 1$  | 2          | $\Lambda_o = \left( \frac{N-p-1}{p} \right) \left( \frac{1-\Lambda^*}{\Lambda^*} \right) \sim F(p, N-p-1)$                         |
| $\geq 1$  | 3          | $\Lambda_o = \left( \frac{N-p-2}{p} \right) \left( \frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F(2p, 2(N-p-2))$       |

• Bartlett의 검정통계량 : N이 충분히 크고, 귀무가설이 참인 조건하에서의 검정통계량에 대한 근사분포 제안

$$\Lambda_o = - \left( N-1 - \frac{p+a}{2} \right) \ln \Lambda^* \sim \chi^2(p(a-1))$$

## • [참고] 다변량 분산분석의 검정통계량 유도

- 분산분석표에서  $SSB_{p \times p} = H$ ,  $SSW_{p \times p} = E$  라고 정의할 때, 다변량 분산분석의 검정통계량은 다음의 식을 만족하는  $\lambda$  의 함수로 표현

$$|H - \lambda E| = 0 \Leftrightarrow |E^{-1}H - \lambda I| = 0$$

- $\lambda$  는  $E^{-1}H$  의 고유값이 되며, 종속(반응)변수가 p개 이면 p개의 고유값이 존재하는 데, '0'가 아닌 고유값을  $\lambda_1 > \lambda_2 > \dots > \lambda_t$  ( $t < p$ ) 로 표현할 때, 다변량 분산분석의 검정통계량은 고유값의 함수로 표현

- p=1인 일변량 분산분석의 경우  $\lambda_1 = \frac{H}{E} = \frac{SSB}{SSW}$

- 다변량 분산분석의 검정통계량

➡ 검정 통계량을 근사 F-분포를 따르는  $\Lambda_0$  로 변환하여 검정

- Wilk's Lambda  $\Lambda^* = \frac{|W|}{|B + W|} = \frac{|E|}{|H + E|} = \prod_{i=1}^t \frac{1}{1 + \lambda_i}$
- Pillai's Trace  $V^* = \text{trace} \left( \frac{W}{B + W} \right) = \text{trace} \left( \frac{E}{H + E} \right) = \sum_{i=1}^t \frac{\lambda_i}{1 + \lambda_i}$
- Hotelling-Lawley Trace  $T^* = \text{trace} (HE^{-1}) = \sum_{i=1}^t \lambda_i$
- Roy's Greatest Root  $R^* = \max \{ \lambda_i \} = \lambda_1$

### ③ p-값 계산과 기각역

- p-값은 종속변수(차원)의 수(p)와 수준의 수(a)에 의해 결정된 근사 F-분포로부터 계산

$$p - \text{value} = \Pr(F(df_1, df_2) > \Lambda_o | H_o)$$

$$\xrightarrow{\text{Rejection region}} \text{if } p - \text{value} < \alpha, \text{ Rejection } H_o$$

$$p - \text{value} = \Pr(\chi^2(p(\mathbf{a} - \mathbf{1})) > \Lambda_o | H_o)$$

$$\xrightarrow{\text{Rejection region}} \text{if } p - \text{value} < \alpha, \text{ Rejection } H_o$$

### ④ 사후분석(다중비교)

- 일반량 분산분석과 동일

## [참고] SAS program

**Proc ANOVA; or Proc GLM;**

Class var1 ; /\*범주형 변수 선언\*/

Model dependent var's = independent var's ; /\* 모형 선언\*/

MANOVA H=factors/PrintH PrintE ; /\*효과 차이 분석 변수\*/

Means factors/Alpha=0.05 post\_method; /\*다중비교 선언\*/

**Run;**

# 예제를 활용한 이해

- 일변량 분산분석 – 독립변수가 종속변수에 미치는 영향을 분석  
(예1) 일변량 분산분석 : 고3 학생들의 영어실력이 지역별 (A시, B시, C시)로 차이가 있는지를 검정
- 다변량분산분석 – 두 개 이상의 종속변수(반응값)의 조합과 독립변수들 사이의 관계를 분석
  - 여러 종속변수의 조합에 대한 효과의 동시검정을 중요시 함  
(예2) 다변량 분산분석 : 고3 학생들의 어학실력(국어, 영어)이 지역별(A시, B시, C시)로 차이가 있는지를 검정

## – MANOVA 기본가정

1. 관측값이 서로 독립이다.
2. 각 집단의 분산-공분산 행렬이 동일하다
3. 모든 종속변수는 다변량 정규분포를 따른다
4. 종속변수들 간의 관계가 선형적이다
5. 종속변수들 간의 상관의 정도가 너무 낮거나 높지 않아야 함

### - 분석목적 :

- 모집단의 중심, 즉 평균벡터 사이에 차이가 있는지의 여부
- 모집단들의 전부 혹은 어떤 것들이 종속벡터변수에 의해 구성된 공간에서 중심이 같은지 혹은 다른지 여부를 조사

# 다변량 분산분석(MANOVA)

<예> 세 가지 교수법에 따라 어휘발달에 차이가 있는지를 검정



P개의 변수가 있고 n개의 집단에 대하여 평균벡터가 같은지 여부 검정시 귀무가설

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \cdots = \begin{bmatrix} \mu_{g1} \\ \mu_{g2} \\ \vdots \\ \mu_{gp} \end{bmatrix}$$

g=2, n=3인 경우

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \end{bmatrix} = \begin{bmatrix} \mu_{31} \\ \mu_{32} \end{bmatrix}$$

$H_1$ : 세 평균 벡터는 모두 반드시 같지는 않다.

# 다변량 분산분석(MANOVA)

<예> 어학실력의 차이 : 고등학교 3학년 학생들의 어학실력을 테스트하기 위하여 서울, 부산, 광주의 세 지역의 학생들을 대상으로 국어와 영어시험을 치르게 하였다. 각 지역에서 10의 학생을 임의로 표본추출하여 시험을 치룬 후에 그 결과를 다음과 같이 정리하였다.

| 서울 |    | 부산 |    | 광주 |    |
|----|----|----|----|----|----|
| 국어 | 영어 | 국어 | 영어 | 국어 | 영어 |
| 82 | 93 | 72 | 58 | 72 | 82 |
| 92 | 65 | 78 | 70 | 68 | 83 |
| 70 | 82 | 83 | 82 | 88 | 80 |
| 77 | 78 | 89 | 75 | 93 | 88 |
| 83 | 90 | 93 | 70 | 92 | 90 |
| 88 | 84 | 88 | 72 | 66 | 68 |
| 95 | 90 | 90 | 63 | 94 | 98 |
| 86 | 88 | 86 | 62 | 88 | 94 |
| 75 | 95 | 90 | 75 | 86 | 92 |
| 90 | 88 | 85 | 53 | 90 | 86 |

$$\bar{Y}_1 = \begin{bmatrix} 83.8 \\ 85.3 \end{bmatrix} \quad \bar{Y}_2 = \begin{bmatrix} 85.4 \\ 68.0 \end{bmatrix} \quad \bar{Y}_3 = \begin{bmatrix} 83.7 \\ 86.1 \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} 84.3 \\ 79.8 \end{bmatrix}$$

$$H_0 : \bar{Y}_1 = \bar{Y}_2 = \bar{Y}_3$$

# 다변량 분산분석(MANOVA)

## <예> 일원 다변량 분산분석표

| 원천  | 제곱합(SS)  | 자유도                    |
|-----|--|------------------------|
| 그룹간 | $SSPB = \sum_i n_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})'$                  | $g - 1$                |
| 그룹내 | $SSPW = \sum_{i=1}^g \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(Y_{ij} - \bar{Y}_i)'$ | $\sum_{i=1}^g n_i - g$ |
| 합계  | $SSPT = \sum_{i=1}^g \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})(Y_{ij} - \bar{Y})'$     | $\sum_{i=1}^g n_i - 1$ |

$H_0$ 의 검정 - Wilks' Lambda [ $\Lambda$ ] 검정

$$\Lambda = \frac{|SSPW|}{|SSPB + SSPW|} = \frac{|SSPW|}{|SSPT|}$$

$$= \frac{\left| \sum \sum (Y_{ij} - \bar{Y}_i)(Y_{ij} - \bar{Y}_i)' \right|}{\left| \sum \sum (Y_{ij} - \bar{Y})(Y_{ij} - \bar{Y})' \right|}$$

➔  $\Lambda$ 가 작으면  $H_0$  기각



# 다변량 분산분석(MANOVA)

## 고유값(eigenvalue)

- 분산분석에서의 집단간 변동과 집단내 변동의 비율을 의미함
- 이 값이 클수록 판별함수의 판별력이 높다고 할 수 있음

## Wilks' Lambda[ $\Lambda$ ]

- 집단내 변동과 총변동(집단간 변동+집단내 변동)의 비율로서 판별 점수의 총변동 중에서 집단간 변동이 설명하지 못한 비율을 나타냄
- 이 값이 작을수록 판별함수의 판별력이 높다고 할 수 있음
- 카이제곱분포에 근접하므로 이를 이용하여 통계적 유의성을 검정할 수 있음
- 유의성 검정은 귀무가설은 “모집단에서의 각 판별함수의 집단중심이 모두 같다”임

# SAS 프로그램

```
data manova_1;  
  do area='seoul', 'puan', 'kwangju';  
    do k=1 to 10;  
      input y1 y2 @@; output;  
    end; end;  
datalines;  
82 93 92 65 70 82 77 78 83 90 88 84 95 90 86 88 75 95 90 88  
72 58 78 70 83 82 89 75 90 70 88 72 90 63 86 62 90 75 85 53  
72 82 68 83 88 80 93 88 92 90 66 68 94 98 88 94 86 92 90 86  
;
```

```
proc GLM; /* proc ANOVA */  
  class area;  
  model y1 y2=area;  
  manova H=area/printh printe;  
  means area/Alpha=0.05 lsd;  
run;
```

# SAS output

- 분산분석
  - Y1(국어) 결과

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 2  | 12.2           | 6.1         | 0.09    | 0.9186 |
| Error           | 27 | 1934.6         | 71.65       |         |        |
| Corrected Total | 29 | 1946.8         |             |         |        |

- Y2(영어) 결과

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 2  | 2091.8         | 1045.9      | 13.82   | <.0001 |
| Error           | 27 | 2043           | 75.67       |         |        |
| Corrected Total | 29 | 4134.8         |             |         |        |

## • 다변량분산분석

- 종속변수의 상관관계

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

| DF = 27 | y1       | y2       |
|---------|----------|----------|
| y1      | 1        | 0.325895 |
| y2      | 0.325895 | 1        |
|         | 0.0906   | 0.0906   |

- 3종 제곱합

H = Type III SSCP Matrix for area

|    | y1     | y2     |
|----|--------|--------|
| y1 | 12.2   | -159.7 |
| y2 | -159.7 | 2091.8 |

- 고유값 : 특성치

Characteristic Roots and Vectors of: E Inverse \* H, where  
H = Type III SSCP Matrix for area  
E = Error SSCP Matrix

| Characteristic Root | Percent | Characteristic Vector V'EV=1 |           |
|---------------------|---------|------------------------------|-----------|
|                     |         | y1                           | y2        |
| 1.2111846           | 100     | -0.0094569                   | 0.0233407 |
| 3.71E-06            | 0       | 0.0221109                    | 0.0016881 |

## – 다변량분산분석

### MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall area Effect

H = Type III SSCP Matrix for area

E = Error SSCP Matrix

S=2 M=-0.5 N=12

| Statistic              | Value     | F Value | Num DF | Den DF | Pr > F |
|------------------------|-----------|---------|--------|--------|--------|
| Wilks' Lambda          | 0.4522446 | 6.33    | 4      | 52     | 0.0003 |
| Pillai's Trace         | 0.5477574 | 5.09    | 4      | 54     | 0.0015 |
| Hotelling-Lawley Trace | 1.2111883 | 7.78    | 4      | 30.19  | 0.0002 |
| Roy's Greatest Root    | 1.2111846 | 16.35   | 2      | 27     | <.0001 |

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

## – 다중비교

Means with the same letter of Y1  
are not significantly different.

| t Grouping | Mean | N  | area  |
|------------|------|----|-------|
| A          | 85.1 | 10 | pusan |
| A          |      |    |       |
| A          | 83.8 | 10 | seoul |
| A          |      |    |       |
| A          | 83.7 | 10 | kwang |

Means with the same letter of Y2  
are not significantly different.

| t Grouping | Mean | N  | area  |
|------------|------|----|-------|
| A          | 86.1 | 10 | kwang |
| A          |      |    |       |
| A          | 85.3 | 10 | seoul |
|            |      |    |       |
| B          | 68   | 10 | pusan |

## [예제 : SPSS 에서의 분석]

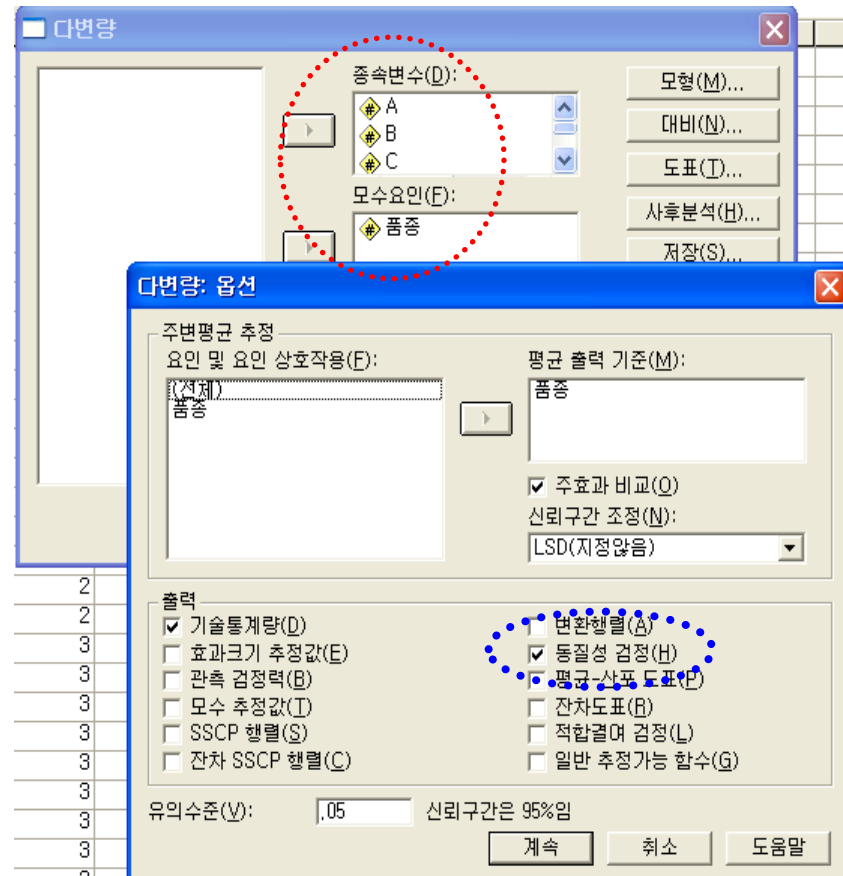
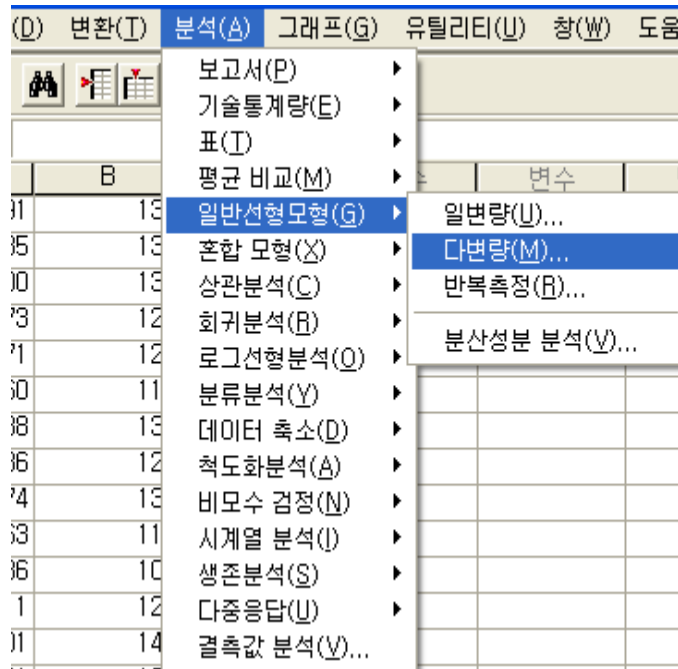
어떤 곤충을 품종별(3종)로 10마리씩 채집하여 부절, 세 부위(A,B,C)의 너비(microns)를 측정. 세 부위의 너비를 종합적으로 볼 때 곤충 품종별로 차이가 있는지 검정

### 1) data 입력

|    | 품종 | A   | B   | C  | 변수 | 변수 |
|----|----|-----|-----|----|----|----|
| 1  | 1  | 191 | 131 | 53 |    |    |
| 2  | 1  | 185 | 134 | 50 |    |    |
| 3  | 1  | 200 | 137 | 52 |    |    |
| 4  | 1  | 173 | 127 | 50 |    |    |
| 5  | 1  | 171 | 128 | 49 |    |    |
| 6  | 1  | 160 | 118 | 47 |    |    |
| 7  | 1  | 188 | 134 | 54 |    |    |
| 8  | 1  | 186 | 129 | 51 |    |    |
| 9  | 1  | 174 | 131 | 52 |    |    |
| 10 | 1  | 163 | 115 | 47 |    |    |
| 11 | 2  | 186 | 108 | 49 |    |    |
| 12 | 2  | 211 | 122 | 49 |    |    |
| 13 | 2  | 201 | 144 | 47 |    |    |
| 14 | 2  | 141 | 131 | 54 |    |    |
| 15 | 2  | 184 | 108 | 43 |    |    |
| 16 | 2  | 211 | 118 | 51 |    |    |

# 다변량 분산분석(MANOVA)

## 2) 분석방법 결정



# 다변량 분산분석(MANOVA)

## 3) 기본가정에 대한 가설검정

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 \quad \left( \begin{bmatrix} \sigma_{1A} \\ \sigma_{1B} \\ \sigma_{1C} \end{bmatrix} = \begin{bmatrix} \sigma_{2A} \\ \sigma_{2B} \\ \sigma_{2C} \end{bmatrix} = \begin{bmatrix} \sigma_{3A} \\ \sigma_{3B} \\ \sigma_{3C} \end{bmatrix} \right)$$

**공분산행렬에 대한 Box의 동일성 검정<sup>a</sup>**

|        |          |
|--------|----------|
| Box의 M | 21,286   |
| F      | 1,483    |
| 자유도1   | 12       |
| 자유도2   | 3532,846 |
| 유의확률   | ,123     |

여러 집단에서 종속변수의 관측 공분산 행렬이 동일한 영가설을 검정합니다.  
a. 계획: Intercept+품종

귀무가설을 기각하지 못하므로  
기본가정에 위배되지 않음

일변량 분산동일성 검정

**오차 분산의 동일성에 대한 Levene의 검정<sup>a</sup>**

|   | F    | 자유도1 | 자유도2 | 유의확률 |
|---|------|------|------|------|
| A | ,457 | 2    | 27   | ,638 |
| B | ,396 | 2    | 27   | ,677 |
| C | ,053 | 2    | 27   | ,949 |

여러 집단에서 종속변수의 오차 분산이 동일한 영가설을 검정합니다.  
a. 계획: Intercept+품종



# 다변량 분산분석(MANOVA)

## 4) 품종별 너비 차이에 대한 가설검정

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \left( \begin{bmatrix} \mu_{1A} \\ \mu_{1B} \\ \mu_{1C} \end{bmatrix} = \begin{bmatrix} \mu_{2A} \\ \mu_{2B} \\ \mu_{2C} \end{bmatrix} = \begin{bmatrix} \mu_{3A} \\ \mu_{3B} \\ \mu_{3C} \end{bmatrix} \right)$$

| 다변량 검정 <sup>c</sup> |                 |         |                       |        |        |      |
|---------------------|-----------------|---------|-----------------------|--------|--------|------|
| 효과                  |                 | 값       | F                     | 가설 자유도 | 오차 자유도 | 유의확률 |
| 절편                  | Pillai의 트레이스    | .997    | 3054.817 <sup>a</sup> | 3,000  | 25,000 | .000 |
|                     | Wilks의 람다       | .003    | 3054.817 <sup>a</sup> | 3,000  | 25,000 | .000 |
|                     | Hotelling의 트레이스 | 366.578 | 3054.817 <sup>a</sup> | 3,000  | 25,000 | .000 |
|                     | Roy의 최대근        | 366.578 | 3054.817 <sup>a</sup> | 3,000  | 25,000 | .000 |
| 품종                  | Pillai의 트레이스    | 1.043   | 9.443                 | 6,000  | 52,000 | .000 |
|                     | Wilks의 람다       | .049    | 29.391 <sup>a</sup>   | 6,000  | 50,000 | .000 |
|                     | Hotelling의 트레이스 | 17.615  | 70.462                | 6,000  | 48,000 | .000 |
|                     | Roy의 최대근        | 17.508  | 151.738 <sup>b</sup>  | 3,000  | 26,000 | .000 |

a. 정확한 통계량  
b. 해당 유의수준에서 하한값을 발생하는 통계량은 F에서 상한값입니다.  
c. 계획: Intercept+품종

귀무가설을 기각하므로  
세 종류의 곤충이  
세 부위의 너비를  
종합적으로 볼 때  
현저한 차이가 존재

# 다변량 분산분석(MANOVA)

## 5) 개별 변수 대한 일변량 가설검정

| 개체-간 효과 검정 |      |                        |     |            |          |      |
|------------|------|------------------------|-----|------------|----------|------|
| 소스         | 종속변수 | 제 III 유형<br>제곱합        | 자유도 | 평균제곱       | F        | 유의확률 |
| 수정 모형      | A    | 25780,200 <sup>a</sup> | 2   | 12890,100  | 64,882   | ,000 |
|            | B    | 209,067 <sup>b</sup>   | 2   | 104,533    | 1,362    | ,273 |
|            | C    | 19,467 <sup>c</sup>    | 2   | 9,733      | 1,116    | ,342 |
| 절편         | A    | 915602,700             | 1   | 915602,700 | 4608,653 | ,000 |
|            | B    | 466253,333             | 1   | 466253,333 | 6076,868 | ,000 |
|            | C    | 75100,033              | 1   | 75100,033  | 8610,195 | ,000 |
| 품종         | A    | 25780,200              | 2   | 12890,100  | 64,882   | ,000 |
|            | B    | 209,067                | 2   | 104,533    | 1,362    | ,273 |
|            | C    | 19,467                 | 2   | 9,733      | 1,116    | ,342 |
| 오차         | A    | 5364,100               | 27  | 198,670    |          |      |
|            | B    | 2071,600               | 27  | 76,726     |          |      |
|            | C    | 235,500                | 27  | 8,722      |          |      |
| 합계         | A    | 946747,000             | 30  |            |          |      |
|            | B    | 468534,000             | 30  |            |          |      |
|            | C    | 75355,000              | 30  |            |          |      |
| 수정 합계      | A    | 31144,300              | 29  |            |          |      |
|            | B    | 2280,667               | 29  |            |          |      |
|            | C    | 254,967                | 29  |            |          |      |

a. R 제곱 = ,828 (수정된 R 제곱 = ,815)  
b. R 제곱 = ,092 (수정된 R 제곱 = ,024)  
c. R 제곱 = ,076 (수정된 R 제곱 = ,008)

A부위는 곤충 품종별 차이가 존재하지만

B, C 부위는 곤충 품종별 차이가 존재하지 않음

채집된 곤충의 품종을 구분하기 위해서는 A 부위의 너비를 측정하여 판단

# 다변량 분산분석(MANOVA)

## 6) 추정된 주변평균

| 추정값  |    |         |       |          |         |
|------|----|---------|-------|----------|---------|
| 종속변수 | 품종 | 평균      | 표준오차  | 95% 신뢰구간 |         |
|      |    |         |       | 하한값      | 상한값     |
| A    | 1  | 179,100 | 4,457 | 169,954  | 188,246 |
|      | 2  | 208,200 | 4,457 | 199,054  | 217,346 |
|      | 3  | 136,800 | 4,457 | 127,654  | 145,946 |
| B    | 1  | 128,400 | 2,770 | 122,717  | 134,083 |
|      | 2  | 122,800 | 2,770 | 117,117  | 128,483 |
|      | 3  | 122,800 | 2,770 | 117,117  | 128,483 |
| C    | 1  | 50,500  | ,934  | 48,584   | 52,416  |
|      | 2  | 48,900  | ,934  | 46,984   | 50,816  |
|      | 3  | 50,700  | ,934  | 48,784   | 52,616  |

| 부위별 너비 |     |    | 품종<br>판단 |
|--------|-----|----|----------|
| A      | B   | C  |          |
| 190    | 143 | 52 | 보류       |
| 174    | 131 | 50 | 1        |
| 211    | 129 | 49 | 2        |
| 218    | 126 | 49 | 2        |
| 130    | 131 | 51 | 3        |
| 138    | 127 | 52 | 3        |

새로  
채집된  
곤충의  
품종  
판단

